

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2001, 13:11:32 ; Search time 3628.34 Seconds
(Without alignments)
11761.720 Million cell updates/sec

Title: US-09-784-340-1

Perfect score: 2759

Sequence: 1 caaccatgcagatcagtgct.....ctgtcagccgttactgacgcg 2759

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_p11:*
13: gb_p12:*
14: gb_p13:*
15: gb_p14:*
16: gb_ba1:*
17: gb_ba2:*
18: gb_ba3:*
19: em_fun:*
20: em_htgo_hum:*
21: em_htgo_inv:*
22: em_htgo_rod:*
23: em_htg_hum1:*
24: em_htg_hum2:*
25: em_htg_hum3:*
26: em_htg_hum4:*
27: em_htg_hum5:*
28: em_htg_hum6:*
29: em_htg_hum7:*
30: em_htg_hum8:*
31: em_htg_inv1:*
32: em_htg_inv2:*
33: em_htg_other:*
34: em_htg_rod:*
35: em_hum1:*
36: em_hum2:*
37: em_hum3:*
38: em_hum4:*
39: em_hum5:*
40: em_hum6:*
41: em_hum7:*
42: em_in:*
43: em_om:*
44: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_p1:*
48: em_p11:*
49: em_p12:*
50: em_p13:*
51: em_p14:*
52: em_p15:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_v13:*
61: gb_v14:*
62: gb_v15:*
63: gb_v16:*
64: gb_v17:*
65: gb_v18:*
66: gb_v19:*
67: gb_v20:*
68: gb_v21:*
69: gb_v22:*
70: gb_v23:*
71: gb_v24:*
72: gb_v25:*
73: gb_v26:*
74: gb_v27:*
75: gb_v28:*
76: gb_v29:*
77: gb_v30:*
78: gb_v31:*
79: gb_v32:*
80: gb_v33:*
81: gb_v34:*
82: gb_v35:*
83: gb_v36:*
84: gb_v37:*
85: gb_v38:*
86: gb_v39:*
87: gb_v40:*
88: gb_v41:*
89: gb_v42:*
90: gb_v43:*
91: gb_v44:*
92: gb_v45:*
93: gb_v46:*
94: gb_v47:*
95: gb_v48:*
96: gb_v49:*
97: gb_v50:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1502	54.4	2786	89	AK025587	AK025587 Homo sapi
2	752	27.3	212904	66	AC021146	AC021146 Homo sapi
3	551	20.0	1650	9	AX040087	AX040087 Sequence
4	362	13.1	212904	66	AC021146	AC021146 Homo sapi
5	72	2.6	138655	72	AC055794	AC055794 Homo sapi
6	72	2.0	1766	93	HS46054	AC006054 Homo sapi
7	56	2.0	169246	62	AC011254	AC011254 Homo sapi
8	50	1.8	386	54	HUM4STS886	L09985 Human Chrom

9	37	1.3	138232	63	AC012582	AC012582 Homo sapi
C	10	1.3	226077	63	AC013296	AC013296 Homo sapi
	11	1.3	2090	97	HSU08854	U08854 Human UDP g
	12	1.3	2114	89	AF180322	AF180322 Homo sapi
	13	1.3	2123	97	HSU06641	U06641 Human UDP g
	14	1.3	181117	87	AC019173	AC019173 Homo sapi
	15	1.2	2075	88	AF072223	AF072223 Macaca fa
	16	1.2	2107	97	HSU59209	U59209 Homo sapien
C	17	1.1	252689	76	AC079433	Mus muscu
	18	1.1	1584	901	AF184901	AF184901 Mus muscu
	19	1.1	1662	89	AF294901	Macaca mu
	20	1.1	2079	93	HSU5162	HSU5162 Homo sapi
	21	1.1	2091	88	AF081793	AF081793 Homo sapi
	22	1.1	2093	97	HSU081793	Y00317 Human mRNA
	23	1.1	2097	88	AF064200	AF064200 Homo sapi
	24	1.1	2108	89	AF112112	AF112112 Macaca fa
	25	1.1	2818	7	OCU27242	OCU27242 Oryctolagus
	26	1.1	120359	85	AC005389	AC005389 Homo sapi
C	27	1.1	134000	92	HS21C080	AL163280 Homo sapi
	28	1.0	1362	10	AX099463	AX099463 Sequence
	29	1.0	1831	95	AX099463	X57565 Rattus norv
C	30	1.0	27229	12	AB019234	AB019234 Arabidops
	31	1.0	130513	85	AC004592	AC004592 Homo sapi
C	32	1.0	138232	63	AC012582	AC012582 Homo sapi
	33	1.0	139516	84	HSJ116913	AL049652 Homo sapi
C	34	1.0	142742	89	AF121782	AF121782 Homo sapi
C	35	1.0	150012	82	AL589868	AL589868 Homo sapi
C	36	1.0	153049	77	AC084734	AC084734 Homo sapi
C	37	1.0	156888	82	AL589847	AL589847 Homo sapi
C	38	1.0	162166	74	AC073477	AC073477 Homo sapi
C	39	1.0	168032	67	AC022785	AC022785 Homo sapi
C	40	1.0	171591	70	AC027018	AC027018 Homo sapi
C	41	1.0	173935	71	AC027797	AC027797 Homo sapi
C	42	1.0	173935	71	AC027797	AC027797 Homo sapi
C	43	1.0	185571	69	AC025800	AC025800 Homo sapi
C	44	1.0	198464	67	AC022764	AC022764 Homo sapi
	45	1.0	226077	63	AC013296	AC013296 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AK025587 2786 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21934 fis, clone HEP04364.
ACCESSION AK025587
VERSION AK025587.1 GI:10438147
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_11b:HEP
Clone:HEP04364.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

REFERENCE

AUTHORS

TITLE

TITLE

JOURNAL

JOURNAL

REFERENCE

REFERENCE

AUTHORS

AUTHORS

TITLE

TITLE

JOURNAL

JOURNAL

REFERENCE

REFERENCE

AUTHORS

AUTHORS

construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

location/Qualifiers
1. 2786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP04364"
/clone_11b="HEP"
/note="cloning vector pME18SF3"
71. 1420
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15179.1"
/db_xref="GI:10438148"
/translation="MPQDRTEENEIEVDLNLVPLGISTWQSVYKLNDFEIRGLTK
MCESEFYNOTLMMKLOETNYDVALIDPVIPCDLMAELLVAFVLTIRISVGNER
SCGKLPAFLSYVPMPTGLDRMTFLERKNSMSLYLFHFIQDYDHFMEERYSKAL
GRPTLCEYVKAFLIIRYVMPPEPPOYPQNPFEVGLHCKPALKPSTTGANRLYDM
SGEDGIVSELSGIFQVTEKANIITASALQIPORVIMRYKGRKSTTGANRLYDM
IPQNDLGHPRKTAFTTHGMNGIYEATIHGVPMVCPITGDDLDNIAHKACGAIVE
INFETWTSIEDLRALRTVITDSSYKEMALSLRIHDQVPLDRAVPIEYPMRKQ
AKHLRSAHDLTWFOHYSIDVYIGFLACVAFALFLFKFLFCSCQFNKTRKLEKE"

BASE COUNT 919 a 463 c 496 g 908 t

ORIGIN

Query Match 54.4%; Score 1502; DB 89; Length 2786;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

202 atgtgctcaactaaagccttctgttaattgactacaggaagccttcgtcatgaaattga 261
|||||
1 ATGTGCTCACTAAAGCCTTCTGTTAATTGACTACAGGAAGCCTTGTGCAATTGTA 60
|||||

262 ggtgtccatcatgacacagacagacagaaatgaatattgttgacactagctct 321
|||||
61 GGTGTCCCATATGCCACAGACAGACAGAAAGAAATGATTTGTTGACCTAGCTCT 120
|||||

322 gaatgtcttgcagagcttaactcgtggaatcagtttaataattgattttttgt 381
|||||
121 GAATGTCTTGCAGAGCTTAATCACTGCAATCAGTTAATAATTAAATGATTTTGT 180
|||||

382 tgaataagaaggaacttaaaatgatgtgagagcttactcaatcagaagcttat 441
|||||
181 TGAATAAGAAGAACTTAAATATATGTGTGAGAGCTTATCTCAATCAGACGCTTAT 240
|||||

442 gaagaagctacaggaacacacactacagatgatgtattagacctgtgacctgtg 501
|||||
241 GAAGAAGCTACAGGAACACCACTACGATGTAATGCTTAAGACCTGTGATTCCTGTGG 300
|||||

502 agagcctgagtgctgagttcttgcagtccttctgtgcacacactagaattctgtag 561
|||||
301 AGAGCCTGAGTGCTGAGTTCTTGCATCCTTTGTGTGCTACACTAGAAATTTGCTAG 360
|||||

562 aggcacatagagcgaagcgtgtggaagccttcagctcacttccatgtatcctgtg 621
|||||
361 AGGCATATAGAGCGAAGCTGTGGAAGCTTCAGCTCCTCACTTCTCATGTACTGTGCC 420
|||||

622 tatgacagagactaacagacagaatgaccccttctggaagaagaataatcattgcttc 681
|||||
421 TATGACAGAGCTAACAGACAGAACTTCTTCTGGAAGAAGTAATAATTCATGCTTTC 480
|||||

682 agtttgcctcacttcgtgattcagagattacagactacattttgtggagagattatag 741
|||||
481 AGTTTGTCTCACTTCTGATTAGAGATTACGACTATTCATTTTGTGGAGAGATTATAG 540
|||||

742 taagcattaggaagccacatcatatgtgagactgtggaagaagcgtgagatatagct 801
|||||
541 TAAGCATTAGGAAGCCCACTACATTATGTGAGACTGTGGAAAGCTGAGATATAGGCT 600
|||||

```

OY      802 aatacgaacatatgttgatlttgaaattcccaacacataccaacctaacttgattgt 861
        |||||||
Db      601 AATACGAACATATATGGATTTTGAATTTCTCTCAACATACCAACCTAATCTTGAATTTGT 660
OY      862 tggaggaattgcactgttaaacctgcgaagcttgctctaaggaatgynaatltgtcca 921
        |||||||
Db      661 TGGAGGATTTGCACCTGTAACCTGCCAAGCTTGGCTTAAGGAATGGAATATTTTGCCA 720
OY      922 gaggtaagggaagatgtagttgtgtgttttctctggggtaactctgtttcaaatgttac 961
        |||||||
Db      721 GAGTTAGGGGGAAGATGATTTGTGTGTTTCTGCGGGCTACCTGTTCAAAATGTTAC 780
OY      982 aagaagaagagcctaataatattgtcttcagccctgcgcgaatcccaagaaggtttatg 1041
        |||||||
Db      781 AGAAGAAAAGGCTAATATATCTTCTTCAGCCCTTGGCCCAAGATCCACAGAGGTGTTATG 840
OY      1042 gaggtacaaaggaanaaaacacatccacattaggagccaactactggtctgtatgtgat 1101
        |||||||
Db      841 GAGGTACAAAGGAAAAAACCATCCACATTTAGAGCCCAATACTCGGCTGTATGATTTGAT 900
OY      1102 accccgaatgatctctctgtgtcaatcccaaaccaagctttatcaatcattgttgat 1161
        |||||||
Db      901 ACCCCGAATGATCTTCTGTGTCATCCCAAAACCAAGCTTTATCATCATGCTGGAAT 960
OY      1162 gaatggatctatgaagcatttaacatgaggtgcctatgtgtggaatgttcccaatttg 1221
        |||||||
Db      961 GAATGGGATCTATGAAGCTATTTACCATGGGGCTTATGTGTGGAGTTCCCATATTTGG 1020
OY      1222 tgaatcagcttgatlaacatagctacacagaagcgaagagcagctgttagaanaaaact 1281
        |||||||
Db      1021 TGATGAGCTTGATATACATGATGCTACATGAAGGCAAGGAGAGCTGTACAAATTAACCT 1080
OY      1282 caaaactatgaacaaggaagattactgtgaggtgttgagaacagtcattacgattctc 1341
        |||||||
Db      1081 CAAAATATGACAAAGGAGGAAATTTACTGAGGGCTTGTGAGAACAGATTAACCATTTCTC 1140
OY      1342 ttataaagaagaatgcatgagattatcaagaattcaacatgataacatgtaagccct 1401
        |||||||
Db      1141 TTATAAAGAAATGCTATGATATCAAGAAATTCACATGATCAACCTGTAAAGCCCT 1200
OY      1402 agatcgaagcagctctctgtgagctgtgtgtcagcgcacaaagaagcgaagcagctgcg 1461
        |||||||
Db      1201 AGATCGAGCAGCTCTCTGTGATGATGCTTGTGATGCGCCACAAAGGAGCCACCTGCG 1260
OY      1462 atcaagctgcccatagacctgaacctggttccaagcactactctatagatgtgtgtcct 1521
        |||||||
Db      1261 ATCAGCTGCCCATGACCTCACTGTTCCAGCACTACTATATGATGTGATTTGGTTCT 1320
OY      1522 gctgacctgtgtggaacgtctatctctgttcacaaatgttttatttcttcgtca 1581
        |||||||
Db      1321 GCTGGCCTGTGTGCGAACCTGCTATTTCTTGTTCACAAAATGTTTTTATTTTCTCTGTCA 1380
OY      1582 aaaaatttaaaactagaaagatagaaaagaggaatagatccttccaaatccaagaa 1641
        |||||||
Db      1381 AAAAATTATTAATAACTAGAAAGATAGAAAAAGAGGAATAGATCTTCCAAATTCAGAAA 1440
OY      1642 gacctgaatgggttaatccgtttaatcccgccacatagaaattgtgtgaaactgtctat 1701
        |||||||
Db      1441 GACCTGATGGGGTAATCCGTTAATTCAGCCCATAGAAATTTGGTGAACCACTTGTCTAT 1500
OY      1702 ttcatatcatctatctgttatttattctttagctatatagccttagaattcca 1754
        |||||||
Db      1501 TTTCAATATATATCTATTTCTGTATTTTATCTTACTATATAGCTATAGCTAAGATTCCA 1553

```

```

RESULT 2
AC021146/c AC021146 212904 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 4 clone RP11-468N14, WORKING DRAFT
DEFINITION SEQUENCE, 24 unordered pieces.
AC021146
AC021146.4 GI:8568861

```

KEYWORDS SOURCE ORGANISM

HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens

REFERENCE

1 (bases 1 to 212904)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS

Waterston, R.H.

TITLE

The sequence of Homo sapiens clone

JOURNAL

Unpublished

COMMENT

2 (bases 1 to 212904)
Waterston, R.H.
Submitted (14-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:7344259.

Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>

Project Information

Center project name: H.NH0468N14
Summary Statistics

Sequencing vector

Sequencing vector: plasmid: 128
Chemistry: Dye-primer ET; 88% of reads
Chemistry: Dye-terminator Big Dye; 12% of reads

Assembly program

Phrap; version 0.990319

Consensus quality

199729 bases at least Q40
Consensus quality: 203731 bases at least Q30
Consensus quality: 206340 bases at least Q20

Insert size

213000; agarose-fp
Insert size: 210604; sum-of-coverage
Quality coverage: 3.75 in Q20 bases; aaarose-fp
Quality coverage: 3.83 in Q20 bases; sum-of-coverage

NOTE

This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will be preserved.

1

1564: contig of 1564 bp in length

1565

1664: gap of unknown length

1665

3085: contig of 1421 bp in length

3086

3185: gap of unknown length

3186

5134: contig of 1949 bp in length

5135

5234: gap of unknown length

5235

8169: contig of 2935 bp in length

8170

8269: gap of unknown length

8270

10341: contig of 2072 bp in length

10342

10441: gap of unknown length

10442

13614: contig of 3173 bp in length

13615

13714: gap of unknown length

13715

17089: contig of 3375 bp in length

17090

17189: gap of unknown length

17190

20701: contig of 3512 bp in length

20702

20801: gap of unknown length

20802

25001: contig of 4200 bp in length

25002

25101: gap of unknown length

25102

29020: contig of 3919 bp in length

29021

29120: gap of unknown length

29121

33356: contig of 4236 bp in length

33357

33456: gap of unknown length

33457

36902: contig of 3446 bp in length

36903

37002: gap of unknown length

42975

42975: contig of 5973 bp in length

42976

43076: gap of unknown length

43076

48339: contig of 5264 bp in length

48340

48439: gap of unknown length

FEATURES	Location/Qualif
*	48430
*	56530
*	56630
*	65213
*	65214
*	65313
*	74715
*	74716
*	74815
*	74816
*	88547
*	88647
*	103367
*	103368
*	103468
*	118168
*	118267
*	132765
*	132766
*	133866
*	154092
*	154093
*	154193
*	157193
*	182144
*	182145
*	182245
*	182245

BASE COUNT ORIGIN	a	c	g	t	others
	69110	36915	37432	67121	2326

Query Match	27.3%	Score 752;	DB 66;	Length 212904;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 752; Conservative	0;	Mismatches	0;	Indels 0;

QY	2	aaccatttgagcatcagtggtgtgagggaaactgcatacgaatgagtcgtacaaagtccattg	61
Db	87449	AACCATTGAGCATCAGTGTGTGAGGGAACTGCCATCATATGAGCTTGACAAAGTCAGCTTGG	87396
QY	62	gtaattctgcctcctgcaagctcctctctgtgttgagctgtggaattctgttggaagtccctggtg	121
Db	87389	GTAATTCTGCTCCTGCAAGCTCTCTGTGTGGCTGTGGAACTTCTGTGGGAAAGTCCCTGGTG	87330
QY	122	tggcccttggaatgaatgacattggtcctaattgtaaglycatctctgaagaagctcatagtg	181
Db	87329	TGGCCCTTGGAATGAATGACATTGGTCTTAATGTCAAAGGTCATTCGTGAAGAGCTCATAGG	87270
QY	182	agagggcgaatgaatgaacagtatatggaactcaaaagcccttgtaattgactcaagaag	241
Db	87269	AGAGGGCAATGAGGTAAACATATTGACTCACTAAAGCCTTCCTTAATTGACTACAGGAG	87210
QY	242	cctcttgcatgaaatttgaagtggtgcacatgycacaagagacaagaacaagaatgaatga	301
Db	87209	CCCTTGCAATTAATAATTGAGGTGGTGCATATGCCACAGACAGAACAGAAATAATGAA	87150
QY	302	acatttggtgacctgacctgtaatgtctctgcagagcttatcaactggaacatcagttata	361
Db	87149	AATATTGTTGACCTTGAAGCTGTGATGATCTTTGCCAGGCTTATCAACCTGGCAATCAATTTATA	87090
QY	362	aaattaaatgattttttgtgtgaataagaggaacttaaaaatgattgtgtgaagcttt	421
Db	87089	AAATTAAATGATTTTTTTGTGTGAATAAGAGGAACCTTTAAAATGATGTGTGAGAGCTTT	87030
QY	422	atctcaatcgaagccttatgaaagaagctacagaacaaactaagatgtaattgcttata	481
Db	87029	ATCTCAATTCAGACGCTTATGAAAGAAAGCTACAGAAACCAACTAGATGTAATGCTTATA	86970
QY	482	gacctgtgattccctgtgagagactgtgtgtagtgccttgcaatccctttgtgtc	541
Db	86969	GACCTGTGATTCCCTGTGAGAGACTGTAGTGCCTAGTTGCTGTGAAGTCCCTTTGTGTCTC	86910
QY	542	aacactgaatctctgttagagagcaatgagagacgtgtggaactccagactcca	601
Db	86909	AACACTGAATTTCTGTAGAGGCAATATGAGAGGAAGCTGTGGAAACTTCCACGCTTCA	86850

QY	602	cttcccaatgacacccgttgcctatgcaggaacaaagaataaactttctgaaaga	661
Db	86849	CTTCCCTATGTAACGTGGCTTATGACAGACATACAGACAGATAACCTTTGTGAAAGA	86790
QY	662	gaaaaaatctaaatctcttcagtttctgtccactcttgatcagatcaactatcat	721
Db	86789	GTAATAAATATCAATGCTTTCACTTTTGTGCACCTCTGTGATTCAGATTACACATATCAT	86730
QY	722	ttttgggaagagtttataagtaaggcattag	753
Db	86729	TTTTGGGAAGAGTTTATATAGTAAGGCATTAG	86698

RESULT	3		PAT
AX040087			18-NOV-2000
LOCUS	AX040087	1650 bp	DNA
DEFINITION	Sequence 10 from Patent WO0063351.		
ACCESSION	AX040087		
VERSION	AX040087.1	GI:11230049	
KEYWORDS			
SOURCE	human.		

BASE COUNT	489 a	330 c	354 g	477 t
ORIGIN				

Query Match	20.0%	Score 551	DB 9	Length 1650
Best local Similarity	100.0%	Pred. No.	4e293	
Matches 551	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1097	tggatccccggaatgagctctctgtgcatcccaaaacgaagctttatcaaccatgct	1156
Db	1100	tggatTACCcCGaATgATCTCTTGGTCATCCAAAACGAACCTTTTATCATCATAGT	1159
QY	1157	ggaatgaatggagatctatgaaagctatttaccatggggtccctatggctggagttcccta	1216
Db	1160	GGAAAGATGGGAGTCATGAAAGCTATTTACCATGGGGTTCCTATGGTGGGAATTCCTATA	1219
QY	1217	tttggctgacgcttgtaataatgctccaatgaagggccaaaggaagcaagctgttagaata	1276
Db	1220	TTTTGGTATCAGCTTGATTAACATAGCTCAATGAAGGCAAGGAGCACTGTATGAATA	1279
QY	1277	aactcaaaactatgaaagcgaagatttactaaggctcttgagaacagcttaaccat	1336
Db	1280	AACCTCAAAACTATGACACAAGGAAGATTACTGAGGGCTTTAGAACACTCTTATCCGAT	1339
QY	1337	tcctcttataaaggaatgctatgagattatacaagaattcaccatgatcaaacctgtaaag	1396
Db	1340	TCTCTTTATAAAGAAATGCTATGAGATTAATCAAGATTTACCAATGATCAACCTGTAAAG	1399
QY	1397	ccctcctatgcgaagcctctctggaatcgagcttgatcatgcccacaaagggccaagcac	1456
Db	1400	CCCCATAGTGGAGCAGTCTTCTGGATCCAGTTTGTCATGCGCCACAAAAGGAGCCCAACAC	1459
QY	1457	ctgagatcagagctgccaatgaaactcaccctggtgttcacagcaactctatagatgttattgg	1516
Db	1460	CTGGATCACTGCTGCCATGACCTCACTGGTTCCAGACACTACTGTATGATGATGTGGG	1519
QY	1517	ttctcgtgacccgtgtggccaactgctataatctgtttccaaaaagtltttatcttc	1576
Db	1520	TTCCTGCTGACCTGTGTGGCACTGCTATATCTGTTCACAAAATGTTTTTATATTTTCC	1579

Qy 1702 ttccatattatctatctgttattatctagctatatagcttagatcca 1754
 Db 34943 TTTCATATATCATCTCTGTATTATTATCTACTATATAGCCTAGATTCCA 34995

RESULT 5
 AC055794 138665 bp DNA HTG 06-JUN-2000
 LOCUS Homo sapiens chromosome 4 clone CTD-2005D20 map 4, WORKING DRAFT
 DEFINITION SEQUENCE, 6 unordered pieces.
 AC055794
 AC055794.2 GI:7801441
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 138665)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 4, clone -2005D20
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 138665)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
 Grand-Plerre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Labroque, K., Lamazares, R., Landers, J., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McCurk, A., McKernan, K., McPheters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tittell, A., Travers, M., Triggillo, J.,
 Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zahoun, J., Zimmer, A. and Zody, M.

Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 14, 2000 this sequence version replaced g1:7382623.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center project name: L7347

Center clone name: 2005_D_20

Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 134168 bases at least Q40

Consensus quality: 136964 bases at least Q30

Consensus quality: 137776 bases at least Q20

Insert size: 176000; agarose-fp

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 5.2 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 4325: contig of 4325 bp in length
 * 4326 4425: gap of 100 bp
 * 4426 14722: contig of 10297 bp in length
 * 14723 14822: gap of 100 bp
 * 14823 30375: contig of 15353 bp in length
 * 30376 30475: gap of 100 bp
 * 30476 57367: contig of 26892 bp in length
 * 57368 57467: gap of 100 bp
 * 57468 95400: contig of 37933 bp in length
 * 95401 95500: gap of 100 bp
 * 95501 138665: contig of 43165 bp in length.
 Location/Qualifiers
 1. 138665
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="4"
 /chromosome="4"
 /clone="CTD-2005D20"
 /clone_11b="CTD Human BAC"
 1. 4325
 /note="assembly-fragment"
 4426. 14722
 /note="assembly-fragment"
 14823. 30375
 /note="assembly-fragment"
 30476. 57367
 /note="assembly-fragment"
 57468. 95400
 /note="assembly-fragment"
 95501. 138665
 /note="assembly-fragment"
 95501="assembly-fragment"
 24794 c 24580 g 43671 t

BASE COUNT 45118 a 24794 c 24580 g 43671 t
 ORIGIN 502 others

Query Match 2.6%; Score 72; DB 72; Length 138665;
 Best Local Similarity 100.0%; Pred. No. 8.9e+28;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2071 ttttcaatttaataagcccttctacataccagcatcaatgactcgaacaagaatt 2130

Db 97061 TTTTTCATTTAAATTAAGCCCTTCAATACCCAGCATTAAGATCTCAGACATGAATT 97120

Qy 2131 gctaaaatgac 2142

Db 97121 GCTAAATAATGAC 97132

RESULT 6
 HSA6054 1766 bp mRNA PRI 16-JUL-1999
 LOCUS Homo sapiens mRNA for UDP glucuronosyltransferase.
 DEFINITION AJ006054.1 GI:4753765
 ACCESSION AJ006054.1 GI:4753765
 VERSION UDP glucuronosyltransferase; ugt2a1 gene.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1766)
 Jedlitschky, G., Cassidy, A.J., Sales, M., Pratt, N. and Burchell, B.
 Cloning and characterization of a novel human olfactory
 UDP-glucuronosyltransferase

Biochem. J. 340 (Pt 3), 837-843 (1999)

JOURNAL MEDLINE 99289328

REFERENCE 2 (bases 1 to 1766)
 Cassidy, A.J.

TITLE Direct Submission
JOURNAL Submitted (07-MAY-1998) Cassidy A.J., Molecular and Cellular Pathology, University of Dundee, University Dept. of Molecular and Cellular Pathology, Ninewells Hospital, Dundee, DD1 9ST, SCOTLAND

FEATURES
 source
 1. .1666
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /rname="factory"
 64. .1647
 /gene="ugt2a1"
 64. .1647
 /gene="ugt2a1"
 /EC_number="2.4.1.17"
 /codon_start=1
 /product="UDP glucuronosyltransferase"
 /protein_id="CAB41974.1"
 /db_xref="GI:4753766"
 /translation="MLNNLLFSLQISLIGTTLGCVNLIMPMEGSHMLNKKIITDELI
 KKEHNVAVASGALFETPTSPSLFEIKVPGKKEIEGVKIDVSTWLENRPS
 TLMRFQEMAKYIKDFHMYSOEICDGLKNOOLMAKLSKEFVYSDVPPGCDIVA
 ILGIDPEMISLAFSPASTYKCKGKRYPPSYPAVLSELTDOMSTDIRNFISHL
 QYMFETLKSNDSTSKALGRPTLCETWGAELWLITVWDFEPRPYLDPFEFVG
 GLCKRPAPLPEKMEEFIQSGKNGVYVSLSMVKNLTKKANLTAALQIPKYL
 WYKGRKPAFLNNTQLEFDIMQNDLGHKTKAFTHGCTNGIYAIHGVMPGV
 MPADPDNIAHMKAGAAVEVNLMTSTVDLSALRTYINEPSYKENARLSLHHDO
 PKRPLDRAVFMIEFVMRHKKAKHLRYAANDLWFOVHSLDIVGLVCYTTAFLVIO
 CCLFSCQKRGKIGKKRRE"

BASE COUNT 542 a 359 c 363 g 502 t

ORIGIN

Query Match 2.0%: Score 56; DB 93; Length 1766;
 Best Local Similarity 100.0%; Pred.No.3.8e-19;
 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1403 gatcgagcagctcttcgagtcgagttgtcatcgccacaaagagcagcact 1458
 |||||||
 Db 1429 GATCGAGCATCTTTCGATCGAGTTGTGTCATCGCCACAAAGAGCCACGACACT 1484

RESULT 7
 AC011254 169246 bp DNA HTG 26-MAY-2000
 LOCUS Homo sapiens chromosome 4 clone RP11-401E5 map 4, WORKING DRAFT
 DEFINITION
 AC011254 12 unordered pieces.
 AC011254.3 GI:7107765
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 169246)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Balwain,J., Barta,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,D.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,D., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission

JOURNAL Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 28, 2000 this sequence version replaced gi:6094610.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIMR
 Web site: http://www.seq.wi.mit.edu/
 Contact: sequence_submissions@genome.wi.mit.edu
 Project information
 Center project name: L1759
 Center clone name: 401.E.5
 Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 129285 bases at least Q40
 Consensus quality: 148859 bases at least Q30
 Consensus quality: 160948 bases at least Q20
 Insert size: 17200; agarose-fp
 Insert size: 168146; sum-of-contigs
 Quality coverage: 3.4 in Q20 bases; agarose-fp
 Quality coverage: 3.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2121: contig of 2121 bp in length
 * 2122 2221: gap of 100 bp
 * 2222 5198: contig of 2977 bp in length
 * 5199 5298: gap of 100 bp
 * 5299 10736: contig of 5438 bp in length
 * 10737 10836: gap of 100 bp
 * 10837 19995: contig of 9159 bp in length
 * 19996 20095: gap of 100 bp
 * 20096 29257: contig of 9162 bp in length
 * 29258 29357: gap of 100 bp
 * 29358 43227: contig of 13870 bp in length
 * 43228 43327: gap of 100 bp
 * 43328 53743: contig of 10416 bp in length
 * 53744 53843: gap of 100 bp
 * 53844 64325: contig of 10482 bp in length
 * 64326 64425: gap of 100 bp
 * 64426 83384: contig of 18959 bp in length
 * 83385 83484: gap of 100 bp
 * 83485 109710: contig of 26226 bp in length
 * 109711 109810: gap of 100 bp
 * 109811 134388: contig of 24578 bp in length
 * 134389 134488: gap of 100 bp
 * 134489 169246: contig of 34758 bp in length.

FEATURES
 source
 1. .169246
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-401E5"
 /clone_lib="RP11 Human Mat: BAC"
 1. .2121
 /note="assembly_fragment"
 2222 5198
 /note="assembly_fragment"
 5299 10736
 /note="assembly_fragment"
 10837 19995
 /note="assembly_fragment"

```

misc_feature      20096..29257
                  /note="assembly_fragment"
misc_feature      29358..43227
                  /note="assembly_fragment"
misc_feature      43328..53743
                  /note="assembly_fragment"
                  clone_end:77
                  vector_side:right"
misc_feature      53844..64325
                  /note="assembly_fragment"
misc_feature      64426..83384
                  /note="assembly_fragment"
misc_feature      83485..109710
                  /note="assembly_fragment"
misc_feature      109811..134388
                  /note="assembly_fragment"
misc_feature      134489..169246
                  /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      52651 a 30209 c 30094 g 55185 t 1107 others
ORIGIN
Query Match      2.0%; Score 56; DB 62; Length 169246;
Best Local Similarity 100.0%; Pred. No. 6.5e-19;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1403 gatcgagcagctctcgtgacgttcgacgttcacgacacgaagcagaacacct 1458
    |||||||
Db 16890 GATCGAGCAGCTCTTCTGATCGATGTTGTCATGCGCCACAAGGACCAAGCAGCAGCT 16945

RESULT 8
LOCUS      HUM4STS886/c 386 bp DNA STS 17-AUG-1993
DEFINITION      Human Chromosome 4 (clone p4-1075) STS4-886, sequence tagged site.
ACCESSION      L09985
VERSION      L09985.1 GI:177711
KEYWORDS      STS; primer; sequence tagged site.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 386)
AUTHORS      Goodid,R.D., disibio,G., Xu,H., Lang,D.B., Daddar,J., Magrane,G.,
Dugalczyk,A., Smith,K.A., Cox,D.R., Masters,S.B. and Myers,R.M.
TITLE      The development of sequence-tagged sites for human chromosome 4
JOURNAL      Hum. Mol. Genet. 2, 1271-1288 (1993)
MEDLINE      94004872
COMMENT      Submitted by: Human Genome Mapping Center,
University of California San Francisco, Box 0925,
San Francisco, CA 94143-0925, USA
e-mail: hgmcp@dbes.cgl.ucsf.edu
Primer A: CATTTAAATCTCTTATTAAGTAC
Primer B: TGCTCTGATGATAAATAGACATC
PCR Components: 25 ng of human genomic DNA, 10 pmol of each
Oligonucleotide,
200 micro-M dNTPs, 0.25 U Taq polymerase (Cetus) in 10 micro-l of
50 mM KCl-20 mM Tris-HCl, pH 8.3 (at room temp), 2.5 mM MgCl-2.
Thermocycler: PE 9600
PCR Profile: Initial denaturation: 94 degrees C for 1.5
minutes
Denaturation: 62 degrees C for 15 seconds Annealing:
72 degrees C for 3.5 minutes PCR Cycles: 30
Final extension: 72 degrees C for 3.5 minutes.
Location/Qualifiers
FEATURES
source      1..386
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
STS      103..295
primer_bind 103..130

```

```

Primer_bind      complement(271..295)
BASE COUNT      134 a 51 c 54 g 144 t 3 others
ORIGIN
Query Match      1.8%; Score 50; DB 54; Length 386;
Best Local Similarity 100.0%; Pred. No. 6.7e-16;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2153 tacactagaatagttgctatattccacataccctccatcat 2202
    |||||||
Db 385 TACACTAGAAATAGTTGCTATATTCCACATACCTCATTAATGATCAT 336

RESULT 9
AC012582      138232 bp DNA HTG 01-JUN-2000
LOCUS      Homo sapiens clone Rp11-3C24, *** SEQUENCING IN PROGRESS ***, 53
DEFINITION      Unordered pieces.
ACCESSION      AC012582
VERSION      AC012582.4 GI:8137462
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 138232)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone Rp11-3C24
Unpublished
TITLE      2 (bases 1 to 138232)
JOURNAL
AUTHORS
REFERENCE
MEDLINE
COMMENT      Direct Submission
Submitted (30-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2000 this sequence version replaced gi:7107861.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 3_C_24
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1640: contig of 1640 bp in length
* 1641 1740: gap of 100 bp
* 1741 2943: contig of 1203 bp in length
* 2944 3043: gap of 100 bp

```


* 3044 4897: contig of 1854 bp in length
* 4898 4997: gap of 100 bp
* 4998 6114: contig of 1117 bp in length
* 6115 6214: gap of 100 bp
* 6215 7665: contig of 1451 bp in length
* 7666 7765: gap of 100 bp
* 7766 9188: contig of 1423 bp in length
* 9189 9288: gap of 100 bp
* 9289 11371: contig of 2083 bp in length
* 11372 11471: gap of 100 bp
* 11472 13194: contig of 1723 bp in length
* 13195 13294: gap of 100 bp
* 13295 14324: contig of 1030 bp in length
* 14325 14424: gap of 100 bp
* 14425 15511: contig of 1087 bp in length
* 15512 15611: gap of 100 bp
* 15612 17362: contig of 1751 bp in length
* 17363 17462: gap of 100 bp
* 17463 19182: contig of 1720 bp in length
* 19183 19282: gap of 100 bp
* 19283 20717: contig of 1435 bp in length
* 20718 20817: gap of 100 bp
* 20818 22066: contig of 1249 bp in length
* 22067 22166: gap of 100 bp
* 22167 23790: contig of 1624 bp in length
* 23791 23890: gap of 100 bp
* 23891 25462: contig of 1572 bp in length
* 25463 25562: gap of 100 bp
* 25563 27362: contig of 2000 bp in length
* 27363 27662: gap of 100 bp
* 27663 29314: contig of 1652 bp in length
* 29315 29414: gap of 100 bp
* 29415 31317: contig of 1903 bp in length
* 31318 31417: gap of 100 bp
* 31418 33429: contig of 2012 bp in length
* 33430 33529: gap of 100 bp
* 33530 35133: contig of 1604 bp in length
* 35134 35233: gap of 100 bp
* 35234 37156: contig of 1923 bp in length
* 37157 37256: gap of 100 bp
* 37257 39661: contig of 2405 bp in length
* 39662 39761: gap of 100 bp
* 39762 41496: contig of 1735 bp in length
* 41497 41596: gap of 100 bp
* 41597 43309: contig of 1713 bp in length
* 43310 43409: gap of 100 bp
* 43410 45445: contig of 2036 bp in length
* 45446 45545: gap of 100 bp
* 45546 47118: contig of 1573 bp in length
* 47119 47218: gap of 100 bp
* 47219 49435: contig of 2217 bp in length
* 49436 49535: gap of 100 bp
* 49536 52546: contig of 3011 bp in length
* 52547 52646: gap of 100 bp
* 52647 53529: contig of 883 bp in length
* 53530 53629: gap of 100 bp
* 53630 55789: contig of 2160 bp in length
* 55790 55889: gap of 100 bp
* 55890 59029: contig of 3140 bp in length
* 59030 59129: gap of 100 bp
* 59130 61713: contig of 2584 bp in length
* 61714 61813: gap of 100 bp
* 61814 63696: contig of 1883 bp in length
* 63697 63796: gap of 100 bp
* 63797 65790: contig of 1994 bp in length
* 65791 65890: gap of 100 bp
* 65891 67845: contig of 1955 bp in length
* 67846 67945: gap of 100 bp
* 67946 69933: contig of 1988 bp in length
* 69934 70033: gap of 100 bp
* 70034 72807: contig of 2774 bp in length
* 72808 72907: gap of 100 bp
* 72908 75815: contig of 2908 bp in length

* 75816 75915: gap of 100 bp
* 75916 79940: contig of 4025 bp in length
* 79941 80040: gap of 100 bp
* 80041 83640: contig of 3600 bp in length
* 83641 83740: gap of 100 bp
* 83741 86921: contig of 3181 bp in length
* 86922 87021: gap of 100 bp
* 87022 91044: contig of 4023 bp in length
* 91045 91144: gap of 100 bp
* 91145 94259: contig of 3115 bp in length
* 94260 94359: gap of 100 bp
* 94360 96910: contig of 2551 bp in length
* 96911 97010: gap of 100 bp
* 97011 100209: contig of 3199 bp in length
* 100210 100309: gap of 100 bp
* 100310 105003: contig of 4694 bp in length
* 105004 105103: gap of 100 bp
* 105104 110031: contig of 4928 bp in length
* 110032 110131: gap of 100 bp
* 110132 115352: contig of 5221 bp in length
* 115353 115452: gap of 100 bp
* 115453 120961: contig of 5509 bp in length
* 120962 121061: gap of 100 bp
* 121062 126155: contig of 5094 bp in length
* 126156 126255: gap of 100 bp
* 126256 130511: contig of 4256 bp in length
* 130512 130611: gap of 100 bp
* 130612 138232: contig of 7621 bp in length.

FEATURES
source
1. .138232
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-3C24"
/clone_1fb="RP11-11 Human Ma.e BAC"
1. .1640
/note="assembly_fragment"
1741. .2943
/note="assembly_fragment"
3044. .4897
/note="assembly_fragment"
4998. .6114
/note="assembly_fragment"
6215. .7665
/note="assembly_fragment"
7766. .9188
/note="assembly_fragment"
9289. .11371
/note="assembly_fragment"
11472. .13194
/note="assembly_fragment"
13295. .14324
/note="assembly_fragment"
14425. .15511
/note="assembly_fragment"
15612. .17362
/note="assembly_fragment"
17463. .19182
/note="assembly_fragment"
19283. .20717
/note="assembly_fragment"
20818. .22066
/note="assembly_fragment"
22167. .23790
/note="assembly_fragment"
23891. .25462
/note="assembly_fragment"
25563. .27362
/note="assembly_fragment"
27663. .29314
/note="assembly_fragment"

Query Match 1.3%; Score 37; DB 6; Length 138232;
Best local Similarity 100.0%; Pred. No. 2.1e-08;

Matches 37: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1124 catcccaaaacgaagctttatcatcgtgtgaa 1160
 DB 110344 CATCCCAAAACCAAGCTTTATCACTCAATGATGAA 110380

RESULT 10
 LOCUS AC013296/C
 DEFINITION Homo sapiens clone Rp11-3M18, LOW-PASS SEQUENCE SAMPLING.
 AC013296.5 GI:9121251
 VERSION AC013296.5
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 226077)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Unpublished
 2 (bases 1 to 226077)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouknight,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Deavrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Gargan,J., Gardyna,S., Grant,G., Hagos,R., Heatford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lebecky,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.,
 Morrow,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrison,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6514003.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L2775
 Center clone name: 3_M18

NOTE: This record contains 246 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 815 914: contig of 814 bp in length
 915 1704: gap of 100 bp
 1705 1804: contig of 790 bp in length
 1805 2602: gap of 100 bp
 2603 2702: contig of 798 bp in length
 2703 3484: gap of 100 bp
 3485 3584: contig of 782 bp in length
 3585 4378: gap of 100 bp
 4378 4378: contig of 794 bp in length

4379 4478: gap of 100 bp
 4479 5257: contig of 779 bp in length
 5258 5357: gap of 100 bp
 5358 6150: contig of 793 bp in length
 6151 6250: gap of 100 bp
 6251 7038: contig of 768 bp in length
 7039 7138: gap of 100 bp
 7139 7940: contig of 802 bp in length
 7941 8040: gap of 100 bp
 8041 8869: contig of 829 bp in length
 8870 8969: gap of 100 bp
 8970 9790: contig of 821 bp in length
 9791 9890: gap of 100 bp
 9891 10695: contig of 805 bp in length
 10696 10795: gap of 100 bp
 10796 11592: contig of 797 bp in length
 11593 11693: gap of 100 bp
 11693 12475: contig of 783 bp in length
 12476 12575: gap of 100 bp
 12576 13364: contig of 789 bp in length
 13365 13464: gap of 100 bp
 13465 14255: contig of 791 bp in length
 14256 14355: gap of 100 bp
 14356 15142: contig of 787 bp in length
 15143 15242: gap of 100 bp
 15243 16023: contig of 781 bp in length
 16024 16123: gap of 100 bp
 16124 16932: contig of 809 bp in length
 16933 17032: gap of 100 bp
 17033 17878: contig of 846 bp in length
 17879 17978: gap of 100 bp
 17979 18786: contig of 808 bp in length
 18787 18886: gap of 100 bp
 18887 19677: contig of 791 bp in length
 19678 19777: gap of 100 bp
 19778 20575: contig of 798 bp in length
 20576 20675: gap of 100 bp
 20676 21463: contig of 788 bp in length
 21464 21563: gap of 100 bp
 21564 22350: contig of 787 bp in length
 22351 22450: gap of 100 bp
 22451 23220: contig of 770 bp in length
 23221 23320: gap of 100 bp
 23321 24108: contig of 788 bp in length
 24109 24208: gap of 100 bp
 24209 24969: contig of 761 bp in length
 24970 25069: gap of 100 bp
 25070 25852: contig of 783 bp in length
 25853 25952: gap of 100 bp
 25953 26726: contig of 774 bp in length
 26727 26826: gap of 100 bp
 26827 27625: contig of 799 bp in length
 27626 27725: gap of 100 bp
 27726 28340: contig of 815 bp in length
 28341 28640: gap of 100 bp
 28641 29413: contig of 773 bp in length
 29414 29513: gap of 100 bp
 29514 30314: contig of 801 bp in length
 30315 30414: gap of 100 bp
 30415 31215: contig of 801 bp in length
 31216 31315: gap of 100 bp
 31316 32099: contig of 784 bp in length
 32100 32199: gap of 100 bp
 32200 32995: contig of 796 bp in length
 32996 33095: gap of 100 bp
 33096 33892: contig of 797 bp in length
 33893 33992: gap of 100 bp
 33993 34782: contig of 790 bp in length
 34783 34882: gap of 100 bp
 34883 35658: contig of 776 bp in length
 35659 35758: gap of 100 bp
 35759 36554: contig of 796 bp in length
 36555 36654: gap of 100 bp

3655 37443: contig of 789 bp in length
37444 37543: gap of 100 bp
37544 38361: contig of 818 bp in length
38362 38461: gap of 100 bp
38462 39228: contig of 767 bp in length
39229 39328: gap of 100 bp
39329 40157: contig of 829 bp in length
40158 40257: gap of 100 bp
40258 41059: contig of 802 bp in length
41060 41159: gap of 100 bp
41160 41959: contig of 800 bp in length
41960 42059: gap of 100 bp
42060 42830: contig of 771 bp in length
42831 42930: gap of 100 bp
42931 43733: contig of 803 bp in length
43734 43833: gap of 100 bp
43834 44657: contig of 824 bp in length
44658 44757: gap of 100 bp
44758 45554: contig of 797 bp in length
45555 45654: gap of 100 bp
45655 46440: contig of 786 bp in length
46441 46540: gap of 100 bp
46541 47356: contig of 816 bp in length
47357 47456: gap of 100 bp
47457 48260: contig of 804 bp in length
48261 48360: gap of 100 bp
48361 49165: contig of 805 bp in length
49166 49265: gap of 100 bp
49266 50058: contig of 793 bp in length
50059 50158: gap of 100 bp
50159 50968: contig of 810 bp in length
50969 51068: gap of 100 bp
51069 51841: contig of 773 bp in length
51842 51941: gap of 100 bp
51942 52748: contig of 807 bp in length
52749 52848: gap of 100 bp
52849 53670: contig of 822 bp in length
53671 53770: gap of 100 bp
53771 54554: contig of 784 bp in length
54555 54654: gap of 100 bp
54655 55442: contig of 788 bp in length
55443 55542: gap of 100 bp
55543 56341: contig of 799 bp in length
56342 56441: gap of 100 bp
56442 57261: contig of 820 bp in length
57262 57361: gap of 100 bp
57362 58170: contig of 809 bp in length
58171 58270: gap of 100 bp
58271 59048: contig of 778 bp in length
59049 59148: gap of 100 bp
59149 59954: contig of 806 bp in length
59955 60054: gap of 100 bp
60055 60858: contig of 804 bp in length
60859 60958: gap of 100 bp
60959 61761: contig of 803 bp in length
61762 61861: gap of 100 bp
61862 62631: contig of 770 bp in length
62632 62731: gap of 100 bp
62732 63524: contig of 793 bp in length
63525 63624: gap of 100 bp
63625 64411: contig of 787 bp in length
64412 64511: gap of 100 bp
64512 65269: contig of 758 bp in length

Query Match 1.3%; Score 37; DB 63; Length 226077;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 catcccaaaccaagctttatcactcatgtgtgaa 1160
|||||
Db 224613 CATCCCAAAACCAAGCTTTATCACTCATGGTGAA 224577

RESULT 11
LOCUS HSU08854 2090 bp mRNA PRI 01-FEB-1995
DEFINITION Human UDP glucuronosyltransferase precursor (UGT2B15) mRNA, complete cds.
ACCESSION U08854
VERSION U08854.1 GI:475758
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2090)
AUTHORS Green,M.D., Ojuru,E.M. and Tepfly,T.R.
TITLE Stable expression of a human liver UDP-glucuronosyltransferase (UGT2B15) with activity toward steroid and xenobiotic substrates
JOURNAL Drug Metab. Dispos. 22 (5), 799-805 (1994)
MEDLINE 95136867
REFERENCE 2 (bases 1 to 2090)
AUTHORS Green,M.D.
TITLE Direct Submission
JOURNAL Submitted (13-APR-1994) Mitchell D. Green, Department of Pharmacology, The University of Iowa, Iowa City, IA 52242, USA
FEATURES
source
1..2090
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HE8a"
/clone_lib="Human liver UniLap cDNA library from Stratagene"
/tissue_type="liver"
22..90
/gene="UGT2B15"
22..1614
/gene="UGT2B15"
/EC_number="2.4.1.17"
/product="UDP glucuronosyltransferase precursor"
/protein_id="AAC50077.1"
/db_xref="GI:475759"
/translation="MSLKWTSVFLILQIS.YFSSGSGKLVMPREYSHWINKTILE ELVONGHEVTVLTSSATPLVNAKSSSAIRLEVPYSLKNDLESLKILDRWIVGYS KNTFWSYQLOELCMEYDYDNKLCDAVNLKMLKLOESKEPVLADALNPGEL LAELNIPFLSLRESVGTPEKNGGFLPPSYVYVAVMSLSDQIMEMERKNKNIHM LYDFNRQIYDLKKMDQFSEVIGRPPLTFEMKGAEMWLIIRTYVDFEPPRLPND FVGLCHCKPAKPLPREMEFVOSSEKNTIVFSLCSIMNSSESAANIASALQIPI KVLWRFEDSKRPPTLGSNTRLYKWLPEONDLIHPKTKAPITGGTIGTAEIYHGIPW GIPLEPADOHNDIAHMKAKGALSYDIRMSSNDILNALSVINPVPYKENVKLSRIH HDQPMKPLDRAVVFIEFYMRHKGAKHLRVAANHNLTWIYHSHDIYAFLLACVAVIFI TKFCLFCEFRKLAKTKKKKKRD"

sig_peptide
CDS
22..1614
/gene="UGT2B15"
/EC_number="2.4.1.17"
/codon_start=1
/product="UDP glucuronosyltransferase precursor"
/protein_id="AAC50077.1"
/db_xref="GI:475759"
/translation="MSLKWTSVFLILQIS.YFSSGSGKLVMPREYSHWINKTILE ELVONGHEVTVLTSSATPLVNAKSSSAIRLEVPYSLKNDLESLKILDRWIVGYS KNTFWSYQLOELCMEYDYDNKLCDAVNLKMLKLOESKEPVLADALNPGEL LAELNIPFLSLRESVGTPEKNGGFLPPSYVYVAVMSLSDQIMEMERKNKNIHM LYDFNRQIYDLKKMDQFSEVIGRPPLTFEMKGAEMWLIIRTYVDFEPPRLPND FVGLCHCKPAKPLPREMEFVOSSEKNTIVFSLCSIMNSSESAANIASALQIPI KVLWRFEDSKRPPTLGSNTRLYKWLPEONDLIHPKTKAPITGGTIGTAEIYHGIPW GIPLEPADOHNDIAHMKAKGALSYDIRMSSNDILNALSVINPVPYKENVKLSRIH HDQPMKPLDRAVVFIEFYMRHKGAKHLRVAANHNLTWIYHSHDIYAFLLACVAVIFI TKFCLFCEFRKLAKTKKKKKRD"
22..1614
/gene="UGT2B15"
91..1611
/gene="UGT2B15"
/EC_number="2.4.1.17"
/product="UDP glucuronosyltransferase"
BASE COUNT 649 a 384 c 435 g 622 t
ORIGIN

Query Match 1.3%; Score 35; DB 97; Length 2090;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1424 gagttgtcattgcgcacaaagaggaagcaacct 1458
|||||
Db 1417 GAGTTGTTCATCGCCACAAAGAGCAAGCACCCT 1453

RESULT 12
LOCUS AF180322 2114 bp mRNA PRI 14-SEP-1999
DEFINITION Homo sapiens UDP-glucuronosyltransferase 2B15 (UGT2B15) mRNA,

UGT2B15-Y85 allele, complete cds.

ACCESSION AF180322.1 GI:5881245

VERSION AF180322.1

KEYWORDS human

SOURCE Homo sapiens

ORGANISM human

REFERENCE 1 (bases 1 to 2114)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Levesque, E., Beaulieu, M., Green, M.D., Tephly, T.R., Belanger, A. and Hum, D.W.

TITLE Isolation and characterization of UGT2B15(Y85): a UDP-glucuronosyltransferase encoded by a polymorphic gene

JOURNAL Pharmacogenetics 7 (4), 317-325 (1997)

MEDLINE 97439504

REFERENCE 2 (bases 1 to 2114)
Levesque, E., Beaulieu, M., Green, M.D., Tephly, T.R., Belanger, A. and Hum, D.W.

AUTHORS

TITLE Direct Submission

JOURNAL Submitted (24-AUG-1999) Molecular Endocrinology, CHUL Research Center, 2705 Laurier Blvd, Ste-Foy, Quebec G1V 4G2, Canada

FEATURES

source

1. .2114
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q13"
1. .2114
/gene="UGT2B15"
/allele="Y85"
39. .1631
/gene="UGT2B15"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B15"
/protein_id="AA055093.1"
/db_xref="GI:5881246"
/translation="MSLKWTSVFLILQSCYFSSGCGVLMVPTREYSHWIMNKITILE ETVVGHEVTVLTSASATLVNASKSAIKLEVPYSLKNTYEDSLKILIDWITGVSKTPFWSYSGLOELCMVEYDYDNKICKDAVILNKIMLKLOESKRPVILADALNPGSEL LAELNIPFLYLSRFSVGTFFKNGGGLFPFSPYVAVSEISDQMIFERKNNIIM IYDFWFOYLDKMDQYSEVLGRPTLFTMGKAEWMLITVWDFEPFPLNPD FVGGILCKPAPKLPKEMEFVOSGSENGIVFSLGSMISNMSESANMIASLAQIPQ KYLMRFEDGKRPNTLGSNTRLYKMLPONDILGHPKTKAFTHGCTNGIYEAIYHGIPMV GILFADONDNTAHMKAGKAALSVDIRMSSRDLLNALKSVINDPYVENWVKLSRIH HDOPKPLDRAVFWIEFVVRHKGAKHLRVAANHLTVIOHSLDVIATFLACAVATVIFI ITRFELCEKRLAKGKKKKRD"

BASE COUNT 670 a 386 c 435 g 623 t

ORIGIN

Query Match 1.3%; Score 35; DB 89; Length 2114;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1424 gagttgtcatgcccacaaaggacgaacacct 1458
|||||
DB 1434 GAGTTGTCTATGCGCCACAAAGGACCAAGCACCCT 1468

RESULT 13

LOCUS HS006641 2123 bp mRNA PRI 16-DEC-1995

DEFINITION Human UDP-glucuronosyltransferase mRNA, partial cds.

ACCESSION U06641

VERSION U06641.1 GI:458398

KEYWORDS human

SOURCE Homo sapiens

ORGANISM human

REFERENCE 1 (bases 1 to 2123)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Irishaid, Y.M. and Tephly, T.R.

TITLE Isolation and purification of two human liver

UDP-glucuronosyltransferases

JOURNAL Mol. Pharmacol. 31, 27-34 (1987)

MEDLINE 87115440

REFERENCE 2 (bases 1 to 2123)
Coffman, B.L., Tephly, T.R., Irishaid, Y.M., Green, M.D., Smith, C., Jackson, M.R., Wooster, R. and Burchell, B.

AUTHORS Characterization and Primary Sequence of a Human Hepatic Microsomal Estriol UDPglucuronosyltransferase

JOURNAL Arch. Biochem. Biophys. 281, 170-175 (1990)

MEDLINE 90343358

REFERENCE 3 (bases 1 to 2123)
Green, M.D., Clarke, D.J., Oturu, E.M., Styczynski, P.B., Jackson, M.R., Burchell, B. and Tephly, T.R.

AUTHORS Cloning and expression of a rat liver phenobarbital-inducible UDP-glucuronosyltransferase (2B12) with specificity for monoterpenoid alcohols

JOURNAL Arch. Biochem. Biophys. 322 (2), 460-468 (1995)

MEDLINE 96032669

REFERENCE 4 (bases 1 to 2123)
Green, M.D.

AUTHORS Direct Submission

TITLE Submitted (07-FEB-1994) Mitchell D. Green, Department of Pharmacology, The University of Iowa, Iowa City, IA 52242, USA

JOURNAL

FEATURES

source

1. .2123
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HLUG4"
/clone_lib="lambda gtl1 human liver cDNA library"
/tissue_type="liver"
<1. .1572
/EC_number="2.4.1.17"
/note="PI 7.4 UDPGT; RLUG4; UGT2B8"
/citation=[1]
/citation=[1]
/codon_start=1
/product="UDP-glucuronosyltransferase"
/protein_id="AA083406.1"
/db_xref="GI:458399"
/translation="VFLLIQLSCYFSSGCGVLMVPTREYSHWIMNKITILEBELVORGH ETVTVSSASTLVNASKSAIKLEVPYSLKNTYEDSLKILIDWITGVSKTPFWSYSGLOELCMVEYDYDNKICKDAVILNKIMLKLOESKRPVILADALNPGSEL FSQLDELQWATYDVSNNKICKDAVILNKIMLKLOESKRPVILADALNPGSEL FVYVRSRISRGVTELEKNGGGLFPFSPYVAVSEISDQMIFERKNNIIM IYDFWFOYLDKMDQYSEVLGRPTLFTMGKAEWMLITVWDFEPFPLNPD FVGGILCKPAPKLPKEMEFVOSGSENGIVFSLGSMISNMSESANMIASLAQIPQ KYLMRFEDGKRPNTLGSNTRLYKMLPONDILGHPKTKAFTHGCTNGIYEAIYHGIPMV GILFADONDNTAHMKAGKAALSVDIRMSSRDLLNALKSVINDPYVENWVKLSRIH HDOPKPLDRAVFWIEFVVRHKGAKHLRVAANHLTVIOHSLDVIATFLACAVATVIFI ITRFELCEKRLAKGKKKKRD"

BASE COUNT 676 a 389 c 427 g 631 t

ORIGIN

mat.peptide

Query Match 1.3%; Score 35; DB 97; Length 2123;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1424 gagttgtcatgcccacaaaggacgaacacct 1458
|||||
DB 1375 GAGTTGTCTATGCGCCACAAAGGACCAAGCACCCT 1409

RESULT 14

LOCUS AC019173 181117 bp DNA PRI 08-NOV-2000

DEFINITION Homo sapiens BAC clone RP11-185H6 from 4, complete sequence.

ACCESSION AC019173

VERSION AC019173.4 GI:11120943

KEYWORDS human

SOURCE Homo sapiens

ORGANISM human

REFERENCE 1 (bases 1 to 181117)
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
99063792
2 (bases 1 to 181117)
Meyer, R., Drone, K., Hawkins, M. and McDill, B.
The sequence of Homo sapiens BAC clone RP11-185H6
Unpublished
JOURNAL
TITLE
AUTHORS
JOURNAL
3 (bases 1 to 181117)
Waterston, R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 181117)
Waterston, R.
Direct Submission
Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gi:7630882.
COMMENT

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

Summary Statistics
Center project name: H_NH0185H06

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pletzer de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://baopac.med.buffalo.edu>)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTD-2005D20. Actual start of
this clone is at base position 1 of RP11-185H6; actual end is at
base position 18117 of RP11-185H6.

FEATURES
source
Location/Qualifiers
1. 181117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-185H6"
/clone_lib="RPCI-11"
repeat_region
1. 71

/rpt_family="ERV1"
271. 444
repeat_region
/rpt_family="ERV.L"
472. 592
repeat_region
/rpt_family="ERV.L"
593. 938
repeat_region
/rpt_family="MaLR"
939. 1217
repeat_region
/rpt_family="ERV.L"
1617. 3178
repeat_region
/rpt_family="L1"
3189. 5532
repeat_region
/rpt_family="L1"
5533. 5818
repeat_region
/rpt_family="Alu"
5819. 6208
repeat_region
/rpt_family="L1"
6219. 6642
repeat_region
/rpt_family="ERV.K"
6660. 7156
repeat_region
/rpt_family="ERV.K"
9336. 10141
repeat_region
/rpt_family="ERV.K"
10149. 10219
repeat_region
/rpt_family="L1"
10242. 10508
repeat_region
/rpt_family="ERV1"
10509. 10978
repeat_region
/rpt_family="L1"
10979. 11006
repeat_region
/rpt_family="ERV1"
10988. 12442
repeat_region
/rpt_family="ERV.K"
12446. 12879
repeat_region
/rpt_family="ERV.K"
12921. 13150
repeat_region
/rpt_family="L1"
13187. 13499
repeat_region
/rpt_family="L1"
13549. 13976
repeat_region
/rpt_family="L1"
14032. 14404
repeat_region
/rpt_family="L1"
14499. 14598
repeat_region
/rpt_family="ERV.L"
14628. 16122
repeat_region
/rpt_family="L1"
16135. 16537
repeat_region
/rpt_family="ERV.L"
16615. 16996
repeat_region
/rpt_family="L1"
16998. 17293
repeat_region
/rpt_family="Alu"
17710. 17766
repeat_region
/rpt_family="Alu"
17801. 18334
repeat_region
/rpt_family="L1"
18343. 18622
repeat_region
/rpt_family="L1"
18623. 19027
repeat_region
/rpt_family="MaLR"
19028. 20370
repeat_region
/rpt_family="L1"
20386. 20698
repeat_region
/rpt_family="ERV1"
20699. 20938
repeat_region
/rpt_family="ERV1"
20977. 21022
repeat_region
/rpt_family="ERV1"
21043. 22020
repeat_region
/rpt_family="ERV1"
22022. 22314
repeat_region
/rpt_family="ERV1"

repeat_region 22321..22476
/rpt_family="L1"
repeat_region 22500..23925
/rpt_family="L1"
repeat_region 24128..24222
/rpt_family="ERV1"
repeat_region 24452..24625
/rpt_family="ERV1"
repeat_region 24645..24899
/rpt_family="L1"
repeat_region 24916..25128
/rpt_family="MALR"
repeat_region 25129..25436
/rpt_family="Alu"
repeat_region 25437..25499
/rpt_family="MALR"
repeat_region 27081..27433
/rpt_family="MALR"
repeat_region 27437..28027
/rpt_family="L1"
repeat_region 28021..28967
/rpt_family="L1"
repeat_region 28954..29769
/rpt_family="L1"
repeat_region 29770..30074
/rpt_family="Alu"
repeat_region 30075..30629
/rpt_family="L1"
repeat_region 30631..31554
/rpt_family="L1"
repeat_region 31675..32162
/rpt_family="ERV1"
repeat_region 32164..32275
/rpt_family="Alu"
repeat_region 32303..32683
/rpt_family="ERV1"
repeat_region 32760..32921
/rpt_family="L1"
repeat_region 32922..33032
/rpt_family="L1"
repeat_region 33033..34170
/rpt_family="L1"
repeat_region 34310..35374
/rpt_family="L1"
repeat_region 35375..36080
/rpt_family="L1"

Query Match 1.3%; Score 35; DB 87; Length 18117;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1424 gagttgtcatgcccacaaagagcgaagcact 1458
|||||
DB 109805 GAGTTGTCTATCGCCACCAAGAGCAGCACCT 109839

RESULT 15
AF072223
LOCUS AF072223 2075 bp mRNA PRI 22-JAN-1999
DEFINITION Macaca fascicularis UDP-glucuronosyltransferase mRNA, complete cds.
ACCESSION AF072223
VERSION AF072223.1 GI:3273888
KEYWORDS
SOURCE crab-eating macaque.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
AUTHORS 1 (bases 1 to 2075)
TITLE Barbier, O., Belanger, A. and Hum, D.W.
Cloning and characterization of a simian
UDP-glucuronosyltransferase enzyme UGT2B20, a novel C19
steroid-conjugating protein

JOURNAL Biochem. J. 337 (Pt 3), 567-574 (1999);
MEDLINE 99112924
REFERENCE 2 (bases 1 to 2075)
AUTHORS Barbier, O., Belanger, A. and Hum, D.W.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-1998) Laboratory of Molecular Endocrinology, CHUL
Research Center, Boulevard Laurier 2700, Quebec, Que G1V 4G2,
Canada

FEATURES
source Location/Qualifiers
1..2075
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/tissue_type="liver; prostate"
14..1606
/EC_number="2.4.1.17"
/note="UGT2B20"
/product="UDP-glucuronosyltransferase"
/protein_id="A008808.1"
/db_xref="GI:3273889"
/translation="MSLKWTSYFLLLQLSCYFSSGCGKVLVMPTEYSHVINKITILE
ELVRRHREYVLTSSASTFVNDKSSAIFFEVYPTSLTKNDMSLMLDLITVTSIS
NSTFLSYFSKLOELCWEYIYSEKLCCKDAVINKLTKETFEVTLADALPCGEL
LAELFNIPIFYSLRFTVGYTFEEKNGGGLFPPSYVYVMSLSDOMTFERRKMTIK
LYDFEWFQIHDIKKMOQFYSEVIGRPRTLLFETMRKADMLIRTYRPFEPRLPVD
FVGLGCKRAKPLPKMEEFVSSGNGVYVTSLSMISMSERANMLASALAOIPQ
KVLWKFDGKRPNTLGSNTRLYKWLPPNDLLGHKRTAFTHTGNTGIVYAIYHGIPVY
GIPLFADQHDNTIVHMKVKGALSLVDIRTHSSRLNALKSVINEPIYKENAMKLSRIH
HDQPMKPLDRAVWFIEFVVRHKGAHLRVAANHLTWIQYHSLDVIALLACVAATFII
ITKCLCFCKRLAKTGKKKKMD"

BASE COUNT 642 a 383 c 415 g 635 t
ORIGIN

Query Match 1.2%; Score 32; DB 80; Length 2075;
Best Local Similarity 100.0%; Pred. No. 7.5e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 ctctctgtcatcccaaaacgaagctttat 1146
|||||
DB 1100 CTCTGTCTATCCCAAAACGCTTTAT 1131

Search completed: August 27, 2001, 15:40:54
Job time: 8962 sec

THIS PAGE BLANK (USPTO)